Prognostic factors for musculoskeletal pain in primary care: a systematic review

Christian D Mallen, George Peat, Elaine Thomas, Kate M Dunn and Peter R Croft

INTRODUCTION
Musculoskeletal pain is a common reason for presenting to primary care where it constitutes up to 18% of a GP’s workload. Although healthcare systems vary between countries, primary care is generally regarded as the point of first contact for musculoskeletal pain, and is the setting in which the majority of cases are assessed and managed.

Estimating the future course of musculoskeletal pain is an important consideration in the primary care consultation for patients and healthcare professionals. An awareness about what is likely to happen in the future allows both patients and professionals to formulate a plan for management. Aside from the individual patient, being able to identify groups at high risk of poor outcome and the factors that may be responsible for the poor prognosis has wider implications for public health initiatives, such as the targeting of obesity as a risk factor for disabling knee pain.

Prospective observational cohort studies play a key role in identifying factors that may help predict future outcome. Research in the field of musculoskeletal pain has largely been organised on the basis of regional pain (for example, low back, neck and shoulder). As a result, studies of prognostic indicators tend to have been viewed in relation to each site separately. An alternative view, however, is that some prognostic indicators may be common across different sites of musculoskeletal pain.

ABSTRACT
Background
Estimating the future course of musculoskeletal pain is an important consideration in the primary care consultation for patients and healthcare professionals. Studies of prognostic indicators tend to have been viewed in relation to each site separately. An alternative view is that some prognostic indicators may be common across different sites of musculoskeletal pain.

Aim
To identify generic prognostic indicators for patients with musculoskeletal pain in primary care.

Design of study
Systematic review.

Setting
Observational cohort studies in primary care.

Method
MEDLINE, EMBASE, PsychINFO and CINAHL electronic databases were searched from inception to April 2006. Inclusion criteria were that the study was a primary care-based cohort, published in English and contained information on prognostic indicators for musculoskeletal conditions.

Results
Forty-five studies were included. Eleven factors, assessed at baseline, were found to be associated with poor outcome at follow up for at least two different regional pain complaints: higher pain severity at baseline, longer pain duration, multiple-site pain, previous pain episodes, anxiety and/or depression, higher somatic perceptions and/or distress, adverse coping strategies, low social support, older age, higher baseline disability, and greater movement restriction.

Conclusion
Despite substantial heterogeneity in the design and analysis of original studies, this review has identified potential generic prognostic indicators that may be useful when assessing any regional musculoskeletal pain complaint. However, it is unclear whether these indicators, used alone, or in combination, can correctly estimate the likely course of individual patients’ problems. Further research is needed, particularly in peripheral joint pain and using assessment methods feasible for routine practice.

Keywords
general practice; musculoskeletal; primary care; prognosis; rheumatology; systematic review.
How this fits in

Prognostic studies of musculoskeletal pain in primary care have typically looked separately at pain at different sites. However, similar prognostic factors for poor outcome appear to emerge irrespective of site. This systematic review has identified 11 ‘generic’ prognostic factors that have been related to poor outcome in studies of pain at different anatomical sites. These are: higher pain severity at baseline, longer pain duration, multiple-site pain, previous pain episodes, anxiety and/or depression, higher somatic perceptions and/or distress, adverse coping strategies, low social support, older age, higher baseline disability, and greater movement restriction. Different study designs and analyses, together with the relative lack of prospective studies of non-spinal pain in primary care, limit the conclusions that can be drawn from this systematic review.

across different sites of musculoskeletal pain: a similar argument has emerged for risk factors for developing new musculoskeletal pain. This has intuitive appeal to the generalist perspective in primary care. It would also make sense, given that individuals with musculoskeletal pain often report pain at more than one site.

In this study, a systematic review of prospective, observational studies in patients presenting to primary care with musculoskeletal pain was conducted. The objective was to identify factors that appear to have prognostic value across different regional pain sites. Indicators of poor outcome for two or more different regional pain complaints (for example, hip and shoulder), are described in this study as being potentially ‘generic’.

METHOD

Participants and methods

Search strategy and study selection. Studies were identified by searching computerised bibliographic databases (MEDLINE, CINAHL, PsychINFO, and EMBASE) from inception to April 2006 using the following keywords: prognosis, course, cohort, predict, epidemiology, statistical models, joint, hand, wrist, shoulder, back, spine, lumbar, neck, cervical, hip, knee, arm, leg, limb, fibromyalgia, pain, ache, arthralgia, general practice, primary care and family physician (full search available on request). Clinicians, epidemiologists, systematic reviewers, and medical librarians were all involved in developing the search strategy. Local experts were contacted to identify additional studies and reference lists of identified studies were checked.

To be included in the review, studies had to be observational, prospective, based in primary care, and include adults consulting with regional (joint or site-specific), or multiple-site musculoskeletal pain. Secondary analyses of randomised controlled trial data were excluded, with only studies published in English considered for inclusion.

All the titles identified were screened with studies not meeting the inclusion criteria excluded. Abstracts and full articles were reviewed to check eligibility, data were extracted, and the quality of all included studies was assessed. The consistency of data extraction and quality assessment on three papers was checked prior to the main extraction.

Data abstraction and quality assessment. The following information was extracted from each of the eligible articles: number of participants, principal outcome measure, definition of poor outcome, (or good outcome when recovery was studied), observed frequency of outcome, prognostic indicators, strength of association, and adjustment for covariates. Only prognostic indicators assessed at baseline or initial consultation were considered. In the event of multiple outcomes (either different outcome domains, such as return to work, pain persistence; or different follow-up points for example, 6 months or 1 year), an association with any was taken as evidence of the indicator’s potential prognostic value. This was true even if multiple outcomes from the same cohort were published in separate articles.

If only crude associations with outcome were presented, significant associations were defined as an unadjusted $P$-value $<0.05$, or odds ratio (OR), or relative risk (RR) with 95% confidence interval (CI) not including 1. If prognostic indicators had been included in a multivariate model, those with an adjusted $P$-value of $<0.05$ or adjusted OR, or RR with 95% CI not including 1, were judged to be significant. If a baseline factor was associated with outcomes at follow up in two or more studies of different regional pain complaints, it was considered to be a possible generic prognostic indicator.

The quality of each study was assessed using a checklist consisting of 15 items, covering aspects related to internal and external validity (Table 1). A search of the general medical literature revealed no single agreed checklist used in previous systematic reviews of observational studies. Instead, common themes were identified from each of the previously used checklists (and those used in previous musculoskeletal systematic reviews), to form the items in the checklist which were discussed and agreed at a consensus meeting. Items were scored positive (+) if present, negative (−) if absent, or unclear (?) if the article did not contain enough information to make an accurate assessment. Disagreements between reviewers were discussed and resolved through consensus.

Data presentation. Results of the quality assessment of individual studies included in the review are
presented separately for each included study. The extracted data is presented separately, and comprises two tables, the first containing the basic details of the studies included in the review: author, date, participants, anatomical site, outcomes, definition of poor prognosis and observed frequencies, where given, (Supplementary Table 1). The second details the association between prognostic indicators and outcome (Table 2). Results of the quality assessment and of the basic study details are available in full in the supplementary information and summarised in this paper.

RESULTS

Selection of studies
A total of 6542 citations were screened (EMBASE and MEDLINE 4230, CINAHL 2012, PsychINFO 300), providing 349 abstracts that appeared to meet the inclusion criteria. Of these 349 abstracts, 276 were excluded (102 not musculoskeletal, 88 randomised controlled trials, 32 duplicates and 54 not conducted in primary care), leaving 73 full text articles. A further 30 studies were excluded after reading the whole article (16 randomised controlled trials, 12 not conducted in primary care and two not relevant). Local experts identified a further two studies, giving a total of 45 included studies (Figure 1).

Quality assessment
Quality assessment scores ranged from 8–14 (mean 11.3; possible range 0–5). One item (E: sample size calculation presented), was not attained by any study whereas two items (A: clearly defined study objective and B: appropriate design for study question), were achieved by all studies (Supplementary Table 2).

Study characteristics
Forty-five studies were included in the systematic review. Approximately half of the studies were of low back pain prognosis (27 studies), with a smaller number relating to spinal pain (six studies), shoulder and/or shoulder-neck pain (four studies), hip pain (two studies), elbow pain (one study), and knee pain (one study). Four studies were of prognosis for general musculoskeletal pain. No primary care-based observational studies of hand or foot pain were identified. Study size varied greatly, ranging from 44 to 3197 participants. Articles included in the review used 14 different outcome domains at follow up. The most frequently used were self-reported disability (22 studies), persistence of symptoms (10 studies), pain severity (10 studies), work absence (eight studies), and patient perceived recovery (eight studies).

Other outcomes included healthcare utilisation (three studies), return to work (three studies), recurrence (two studies), medication use (two studies), bed rest, catastrophising, pain-related fear, quality of life, and hip replacement (all used as outcomes in one study). Studies using the same outcome domain (for example, disability), frequently used different outcome measures (for example, for low back pain studies including disability as an outcome domain, three different measures were used). The length of time participants were followed up for ranged from 2 weeks to 5 years.

Table 1. Quality assessment checklist.

<table>
<thead>
<tr>
<th>Criteria used to assess study</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Clearly defined study objective</td>
</tr>
<tr>
<td>B. Appropriate design for study question</td>
</tr>
<tr>
<td>C. Inclusion and exclusion criteria clear and appropriate</td>
</tr>
<tr>
<td>D. Representative sample (and comparison)</td>
</tr>
<tr>
<td>E. Sample size calculation presented</td>
</tr>
<tr>
<td>F. Appropriate selection of outcome</td>
</tr>
<tr>
<td>G. Appropriate measurement of outcome</td>
</tr>
<tr>
<td>H. Standardised collection of data</td>
</tr>
<tr>
<td>I. Adequate length of follow up for research question</td>
</tr>
<tr>
<td>J. Baseline participation &gt;70% (all groups)</td>
</tr>
<tr>
<td>K. Losses and drop outs &lt;20%</td>
</tr>
<tr>
<td>L. Adequate description of losses and completers</td>
</tr>
<tr>
<td>M. Appropriate analysis of outcomes measured</td>
</tr>
<tr>
<td>N. Numerical description of important outcomes given</td>
</tr>
<tr>
<td>O. Adjusted and unadjusted calculations provided (with CI if appropriate)</td>
</tr>
</tbody>
</table>

CI = confidence interval

Figure 1. Results of systematic search and selection of studies.

Figure 1. Results of systematic search and selection of studies.
Prognostic indicators

In total, 63 different prognostic indicators were identified, which had been investigated in at least one study and shown to be associated with outcome. Eleven of these prognostic indicators were associated with outcome in at least two different studies of pain at different anatomical sites, although low back pain and spinal pain were presented separately, they were regarded as one site for the purposes of this analysis because of the overlap between them; supplementary Table 1).

The other 49 prognostic indicators were associated with outcome in at least one study, but only for a single anatomical site. Of the eleven indicators at baseline that were associated with poor outcome across more than one anatomical site, four were pain characteristics (higher pain intensity, longer pain duration, multiple-site or widespread pain and previous episode of pain), and two were related to psychological conditions (higher levels of anxiety or depression and higher somatic perceptions or psychological distress).

Other indicators identified as being associated with a poor prognosis were movement restriction, adverse coping strategies, low levels of social support, and older age. Higher baseline pain intensity and disability were associated with poor outcome at all of the anatomical sites in this review (general musculoskeletal, spinal, low back, shoulder-neck, hip, knee, and elbow). Longer pain duration at baseline was indicative of poor prognosis for spinal pain, low back pain, shoulder pain, and hip pain, while the presence of widespread pain was a prognostic indicator for poor outcome in spinal pain, low back pain, and shoulder pain. To assess whether study quality was associated with the prognostic indicators identified, the analysis was repeated after excluding 11 studies with quality (Supplementary Table 1).

Table 2. Association between prognostic indicators and outcome, crude, and after adjusting for covariates (bold).

<table>
<thead>
<tr>
<th>Prognostic indicator at baseline</th>
<th>Association with poor outcome</th>
<th>Low back (n = 27)</th>
<th>Spinal (n = 6)</th>
<th>Shoulder/neck (n = 4)</th>
<th>Hip (n = 2)</th>
<th>General musculoskeletal (n = 4)</th>
<th>Knee (n = 1)</th>
<th>Elbow (n = 1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher pain intensity Positive</td>
<td>[21][27][35][44][50][52][56]</td>
<td>[19][24][26][45][48]</td>
<td>[25][47][55]</td>
<td>[39]</td>
<td>[16][37]</td>
<td>[53]</td>
<td>[46]</td>
<td></td>
</tr>
<tr>
<td>Longer pain duration Positive</td>
<td>[12][13][4][21][27][29]</td>
<td>[30][32][42][44][50][56]</td>
<td>[20][31][33][35][49]</td>
<td>[19][24][26]</td>
<td>[23][25]</td>
<td>[47][55]</td>
<td>[39][51]</td>
<td>[53]</td>
</tr>
<tr>
<td>Multiple site / widespread pain Positive</td>
<td>[21][29][32][35][44]</td>
<td>[26]</td>
<td>[25][47][55]</td>
<td>[36]</td>
<td>[46]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous episode Positive</td>
<td>[14][20][33][32][52]</td>
<td>[31][42][49][56]</td>
<td>[45]</td>
<td>[23][47][55]</td>
<td>[53]</td>
<td>[46]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greater movement restriction Positive</td>
<td>[32][40]</td>
<td>[20][31]</td>
<td>[23]</td>
<td>[39]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Higher disability Positive</td>
<td>[13][14][20][21][31]</td>
<td>[33][35][42][44][50]</td>
<td>[24][26]</td>
<td>[23][47]</td>
<td>[39]</td>
<td>[37]</td>
<td>[53]</td>
<td>[46]</td>
</tr>
<tr>
<td>Higher anxiety / depression Positive</td>
<td>[17][21][30][43][49]</td>
<td>[35]</td>
<td>[24][26][45]</td>
<td>[18][34][36]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Higher somatic perceptions/distress Positive</td>
<td>[21][49]</td>
<td>[26][45]</td>
<td>[55]</td>
<td>[33]</td>
<td>[53]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coping strategies Positive</td>
<td>[21][43][54]</td>
<td>[26]</td>
<td>[47]</td>
<td>[18]</td>
<td>[46]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social support Positive</td>
<td>[23][44][49][52][56]</td>
<td>[19]</td>
<td>[47]</td>
<td>[55]</td>
<td>[51]</td>
<td>[36]</td>
<td>[53]</td>
<td>[46]</td>
</tr>
</tbody>
</table>

Papers 14 and 15 are from the same study and have only been included in Table 2 once (as study 14). Papers 21 and 41 are from the same study and have only been included in Table 2 once (as study 21). Bold indicates adjusted for covariates. *Predictor of becoming lower back pain free. †Prognostic factor for recovery. ‡Predictors of early improvement.
This did not alter the pattern of findings on generic prognostic indicators. Heterogeneity in either study populations, outcomes, the specific measurement of each prognostic indicator, or approaches to statistical analysis, prevent any meaningful pooled quantitative estimate of the association between prognostic indicator and poor outcome. For example, in eight of the studies reporting a positive association between higher pain intensity and outcomes at 1 year, two provided unadjusted $P$-values $<0.05$; two provided adjusted OR $= 8.1$; no CI; two provided adjusted RR $= 0.93$, 95% CI $= 1.2$ to 3.3; two provided unadjusted RR $= 0.98$; two provided adjusted RR $= 2.8$, 95% CI $= 1.8$ to 4.2; one provided adjusted hazard ratios for time to recovery, 0.99, 95% CI $= 0.99$ to 1.00; and one provided a percentage of variance in outcome explained by pain intensity within a multivariate model (5%).

**DISCUSSION**

**Summary of main findings**

This is the first systematic review to compare prognostic factors for musculoskeletal pain in primary care across a number of different anatomical sites. Despite the high degree of heterogeneity in study population, design, and analysis, certain generic prognostic factors do consistently emerge. Pain characteristics at the time of consultation (intensity, duration, previous episodes and multiple sites), levels of disability and psychological factors such as anxiety or depression, and higher levels of psychological distress, were all associated with subsequent outcome in multiple studies investigating different pain sites. Other generic prognostic indicators identified by the review included greater movement restriction, older age, lower levels of social support, and adverse coping strategies, although these indicators were identified by fewer original studies. The implication of this review is that when estimating the strength of association (for example, the OR) between a prognostic indicator and outcome is inadequate for describing how well it correctly classifies individuals, particularly if the indicator is used in isolation rather than in combination with others. For example, in one study high disability at the point of initial consultation simply estimating the strength of association may include prognostic indicators investigated only previously within studies of spinal pain, such as some psychosocial and work related factors. In the assessment of methodological quality, a 15-point checklist was used, allocating equal weighting to each of the items assessed. Studies for this review were excluded on the basis of quality assessment scores. Removing ‘lower quality’ studies did not affect the overall conclusions. The use of such tools is controversial and, in the absence of empirical evidence to support the concept of ‘quality’ in a descriptive-prognostic study, the assumption behind equal weighting of checklists is nevertheless questionable. It is possible for studies that are fundamentally flawed in one crucial aspect to still achieve the same high rating as a fundamentally sound study that fails a minor criterion of quality.

Publication bias or underreporting in published studies would potentially provide more instances where generic indicators were not associated with poor outcome, tending to weaken support for their prognostic value. The search strategy for this study included several bibliographic databases from the time of their inception until April 2006. In an effort to identify studies that had not been published within the search period, local experts were contacted, resulting in two further studies, one of which has subsequently been published, and the other is a chapter in a doctoral thesis.

Also included were studies irrespective of their methodological quality, although omitting those with lower quality assessment scores did not substantially alter the conclusions. In common with other reviews, the search did not cover further any grey literature, and was restricted to studies published in English. Although publication bias commonly occurs, it does not always affect the validity of a review’s conclusions, and there is little evidence that not including non-English language research results in biased estimates.

**Comparison with existing literature**

This review does not make clear if these generic indicators would help primary care clinicians arrive at the correct prognosis for an individual patient at the point of initial consultation. Simply estimating the strength of association (for example, the OR) between a prognostic indicator and outcome is inadequate for describing how well it correctly classifies individuals, particularly if the indicator is used in isolation rather than in combination with others. For example, in one study high disability at
initial consultation had the strongest association with poor outcome 3 months later.\textsuperscript{10} Nevertheless, three-quarters of those with high disability at the initial consultation had recovered 3 months later.

A small number of studies in this review reported how often multivariate models correctly classified individuals with poor outcome. In predicting which patients would be put on the waiting list for hip replacement in the 4 years after initial consultation, Birrell \textit{et al.}\textsuperscript{20} found that a model including baseline pain intensity in combination with radiographic severity and use of a walking stick attained sensitivity ranging from 24.2\% to 100\% and specificity from 32.1\% to 100\%.

Burton \textit{et al.}\textsuperscript{11} used a model including pain duration, pain intensity, somatic perceptions with pain coping, and locus of control to correctly identify 72\% of low back pain patients recovered at baseline pain intensity, pain persistence and contained generic prognostic factors (for example, site-specific indicators, it may still be difficult to provide individual patients with an accurate specificity from 32.1\% to 100\%.

\textbf{Implications for research and clinical practice}

Why should similar factors be consistently associated with outcome across several seemingly different regional pain complaints, and despite often considerable heterogeneity between studies? One suggestion is that, in very broad terms at least, different regional pain syndromes share similar underlying attributes and clinical courses. Most are recurrent, the likelihood of future episodes is higher in those with a previous history of attacks, and the longer pain and disability are established, the more likely they are to persist. There will be a degree of auto-correlation in measures of pain and disability severity over time, such that individuals with high levels at one point in time are more likely to have high levels at a second point in time.

Why psychosocial factors such as anxiety, depression, coping, and social support should be associated with poor outcome across different regional pain syndromes is more difficult to explain. The explicit use of underlying theoretical models has typically been lacking.\textsuperscript{40} The distinction between different regional pain syndromes may be more apparent than real. In every study, a proportion of the participants will have multiple-site pain and these individuals will be common to each study, perhaps contributing to similar findings. Irrespective of the reasons for similar findings across different studies of different regional pain syndromes, this review highlights the potential of exploiting this. This review has identified factors that appear to be of prognostic value irrespective of the site of pain. Implicit in this is the potential for using a generic approach as the foundation for predicting poor outcome among patients consulting primary care with musculoskeletal pain. Further research is warranted, particularly in regard to predicting the future course of non-spinal pain. Part of the remit of this research would be to investigate the extent to which these generic prognostic indicators sufficiently discriminate between patients who do and do not go on to experience a poor outcome following primary care consultation.

\textbf{Supplementary information}

Additional information accompanies this paper at http://www.rcgp.org.uk/bjgp-supinfo

\textbf{Funding body}

Christian Mallen is funded by an Arthritis Research Campaign Primary Care Fellowship (Reference number MO669)

\textbf{Competing interests}

The authors have stated that there are none

\textbf{Acknowledgements}

We would like to thank Jo Jordan (systematic reviewer) and Steve Sharp (health librarian) for their help with the review and search strategy.

\textbf{REFERENCES}


